The Informal Sector Wage Gap: New Evidence using Quantile Estimations on Panel Data

Olivier Bargain, Prudence Kwenda

Working Papers / Documents de travail

The Informal Sector Wage Gap: New Evidence using Quantile Estimations on Panel Data

Olivier Bargain
Prudence Kwenda

WP 2013 - Nr 60

The Informal Sector Wage Gap: New Evidence using Quantile Estimations on Panel Data*

Olivier Bargain and Prudence Kwenda

Submitted: Jan. 2010; revised: Dec. 2012; accepted: June 2013
Forthcoming in *Economic Development and Cultural Change*

Abstract

We estimate the informal-formal sector pay gap throughout the conditional wage distribution using panel data from Brazil, Mexico and South Africa. We control for time-invariant unobservables and identification is stemming from inter-sector movers. We control for observables in a non-linear way using propensity score reweighting and carefully check for potential measurement errors. Using similar definitions of informality, we obtain consistent results for all three countries: Informally employed workers earn much less than formal workers primarily because of lower observable and unobservable skills. Estimates of the conditional wage gap show that they are also underpaid compared to their formal sector counterparts. In all three countries, the informal wage penalty is larger in the lower part of the conditional distribution and tends to disappear at the top, i.e., the informal sector increases wage dispersion. The magnitudes of these effects vary across countries, with the largest penalties in lower conditional quantiles of South Africa and more modest wage gaps in Latin America. We suggest explanations in line with different legal and labor market conditions.

Key Words : wage gap, informal sector, quantile regression, fixed effects, propensity score, conditional random effects.

***Acknowledgements** : Bargain is affiliated with Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS, and IZA. Kwenda is affiliated with the University of Witwatersrand, Johannesburg. We are grateful to Blaise Melly, Eliane Badaoui, Bill Maloney, Martin Rama, Eric Strobel, Arthur van Soest, Elizabeth Villagomez and participants/discussants of the 2009 RIW conference, workshops at IZA and Brunel University and seminars at CEPS-INSTEAD and UCD for useful advices. The usual disclaimers apply. Correspondence to: Olivier Bargain, IZA, Schaumburg-Lippe Str. 5-7, 53113, Bonn, Germany. Email: olivier.bargain@univ-amu.fr

JEL Classification : J21, J23, J24, J31, C14, O17

1 Introduction

Most developing and emerging economies are characterized by the presence of informal employment. While many definitions exist, an informal, unregulated labor market can be seen as one where workers are unregistered and not liable to taxes and contribution, not subject to labor market regulations and excluded from social security coverage (pension, benefits) or the right to a minimum wage.[1] Such type of employment represents a large share of the working force worldwide and a long-term feature of the labor markets in developing countries, particularly in Africa and Latin America (see the different studies referenced in Perry et al., 2006, ed.). While it is often suspected to cumulate low earnings, bad work conditions and poor employment benefits, evidence is mixed mainly because of the highly heterogeneous nature of informal employment. The existence of an informal sector must have crucial implications on the earnings structure, on the functioning of labor markets overall and, ultimately, on the policies that should be adopted by governments to maximize the welfare of a nation. The present paper aims to address the first point, namely to improve the measurement of earnings distributions in emerging economies and, more precisely, to provide a distributional analysis of the informal-formal pay gap – an important aspect to understanding wage structures and the notion of decent work in developing countries.

Several reasons can explain why workers in undeclared and uncovered jobs are paid differently from identical, formally employed workers. An informal wage penalty may arise if labor market regulation (minimum wages, higher unionization) not only keeps a large part of the labor force out of formal employment but also pushes up formal sector wages above market-clearing levels. It may also derive from lower bargaining power among informally employed workers (Carneiro and Henley, 1998). These situations often – but not always – characterize informal workers in firms which are themselves unregistered. In that case, an informal wage penalty may also be the result of a firm size effect. Indeed, larger firms pay more and, at the same time, are more likely to be formal because of larger exposure to the risk of being caught defaulting (see Badaoui et al., 2010). Finally, informal sector wage penalties may reflect compensating differentials if non-pecuniary amenities are attached to informal employment, such as more flexible hours, training, and tax savings.[2] Inversely, informal jobs could command higher earnings to compensate for the value of lost fringe

[1] Note that we focus on informal versus formal salary work in the present paper (we extend the comparison to self-employment in Bargain and Kwenda, 2011). For this reason, one could argue that "informal employment" rather than "informal sector" may be a more accurate term. We use both expression indifferently, the latter being consistent with the general terminology in the literature.

[2] For low-skilled youth and older workers, informal salaried jobs are sometimes described as offering an entry point to the labor market that partially allows them to remedy deficient schooling or the obsolescence

benefits such as medical coverage and old age pension (net of payroll contributions), unless workers do not value these benefits.[3] Informal wages could also be higher (or the informal sector penalty be reduced) if employers have to pay some taxes or contribution before paying a formal wage.

This range of possible explanations calls for an appropriate characterization of the informal sector wage gap. Estimations of the mean gap, as suggested in several studies and for several countries (e.g., Badaoui et al., 2008, for South Africa), conceal the variety of situations that may exist. Instead, the present paper suggests using quantile regression (QR) to assess informal wage penalties at different points along the conditional wage distribution. Since workers may sort into formal and informal jobs, acting on the returns to their skills and competencies in each sector, we run fixed effects quantile regressions (FE-QR) to control for workers' (time-invariant) unobserved characteristics. The informal-formal sector wage gap is identified on cross-sector movers. For this reason, we particularly focus on potential measurement errors and provide extensive robustness checks. Finally, it is important to interpret our estimates, obtained by regressions on the pooled sample of informal and formal workers, as *conditional* wage differentials between sectors. That is, controlling for the distributions of workers' characteristics across sectors (the "between" effect), we focus on the potentially different pay settings between sectors for otherwise identical workers (the "within" effect). Since it may be restrictive to control linearly for workers' characteristics, we also improve the comparison by adjusting our estimation using matching methods. The robust combination of fixed-effect estimation and propensity score weighting is originally extended to quantile estimations.

Importantly, we replicate our estimations for three different countries. Previous studies usually consider one country at a time, using specific methods and definitions of informality to identify the conditional sector wage gap. We suggest applying a uniform estimation and identification strategy as well as a very comparable definition of informality to three emerging economies which have received much attention in the literature, namely South Africa, Mexico and Brazil. This harmonized empirical approach should contribute making the finds more generalizeable to labor markets in developing countries in general.[4] In

of skills through on-the-job training unavailable to them in formal salaried jobs. For women, the balance between work and family responsibilities may render the greater flexibility and autonomy of informal jobs a better match.

[3]This can happen for many reasons: a lack of information about benefits and the functioning of social security programs, the fact that these services are universally provided or traditionally provided through family support, or the fact that workers are aware of inefficiencies in formal social protection.

[4]It is also interesting to replicate the exercise for countries with different data quality. The small South African sample, used in other contributions (e.g., Badaoui et al., 2008), yields relatively less precise

fact, despite different degree of informality between these countries, and contrasted labor market history and legal institutions, results point to consistent patterns over all three countries: First, salary workers in the informal sector are underpaid compared to their formal sector counterparts, yet the penalty is small for a majority of workers. Second, the penalty is smaller after controlling for fixed effects, i.e., the very large unconditional wage gap is not only explained by 'better' observed characteristics in the formal sector but also by better unobserved skills. This means that pooled cross-sectional estimates of wage differentials greatly overstate the actual wage penalties suffered by informal workers. Third, quantile estimations unveil another consistent pattern, namely that informal wage penalties are significant in the lower part of the conditional distribution while they tend to disappear at the top. In other words, those who do badly conditional on their observed characteristics do especially poorly in informal salary work. That the informal sector increases wage dispersion reflects the heterogeneity of workers/jobs in this sector – it is also consistent with the wage compression that can be expected in the regulated sector.

Beyond qualitatively comparable patterns, we find differences in the magnitude of these effects between countries, which are in line with different institutional and legal backgrounds. The largest informal wage penalties are found in the lower half of the conditional wage distribution in South Africa. This is suggestive of the fact that legal advantages in formal employment (e.g., unionization) are more effective in this country, resulting in workers left out of formal jobs and/or possibly stronger bargaining power in the formal sector. These explanatory factors do not apply uniformly over the wage distribution, however. Indeed, a half of the informal sector resembles that in the two Latin American countries, characterized by more modest wage gaps overall. Higher employer costs attached to formal employment in Brazil and Mexico may simultaneously explain the large extent of informal work and the relatively smaller sector wage gap in these countries, as firms possibly recoup high employers' payroll taxes paid to hire formal workers. Finally, we suggest that informal wage penalties may only partly be related to the firm size effect, especially at the top of the conditional distribution and more frequently in Brazil, where many informal workers are to be found in large formal firms.

The paper is organized as follows. Section 2 positions the paper in the literature, presents the evaluation problem and provides some background information. Section 3 describes the data and the construction of raw wage differences after accounting for taxes in the

estimations than for other countries. Results for this country can nonetheless be compared and reconciled with the more robust results on the large Mexican data. We also address the potential problem of using short panels by replicating our results for different panel durations on Mexican data and by comparing our baseline with estimators which do not suffer from the incidental parameter problem.

formal sector. The econometric approach is detailed in section 4. In Section 5, we present and discuss the empirical results and several robustness checks. Section 6 concludes.

2 Background

We first position our contribution in the literature on informal-formal sector wage gaps. We discuss the evaluation problem at stake in the paper. Finally, we explain our definition of informal employment and provide some background information on labor markets in the three countries under study.

2.1 Literature and Contribution

Several previous studies have estimated the conditional wage gap between informal and formal sectors. We provide several references for Brazil, South Africa and Mexico and other countries throughout this section (we do not aim at an exhaustive survey but simply cite some of the studies which provide relevant comparison points). They are summarized in Table 1, including the different estimation techniques and the definition of informality.[5] Most studies are affected by at least one of the three following shortcomings that we simultaneously address in our empirical approach.

First, most studies focus on comparisons at the mean. This necessarily conceals important information and may explain – in addition to the use of different methodologies and sample selection – the difficulty to reconcile the different estimates obtained in the literature. For instance, some studies find a large informal wage penalty (e.g., Funkhouser, 1997, for El Salvador, Gong and Van Soest, 2002, for Mexico), even after controlling for workers' heterogeneity, while others show that this penalty tends to disappear in that case (e.g., Pratap and Quintin, 2006, for Argentina, Badaoui et al., 2008, for South Africa). Ignoring distributional issues is also a strong limitation, given the intricate question of how informality affects earnings inequality. Notice that a few studies make use of quantile estimations to estimate sector wage gaps along the conditional wage distribution. Most of them ignore the problem of selection, however, and do not attempt to control for unobservables (e.g., in Perry et al., 2006, ed.). Rare exceptions exist and essentially adopt quantile regressions corrected for selection using instrumental variables, as in the application of Tannuri-Pianto and Pianto (2008) for Brazil. We compare our results to theirs in what follows.

[5] The lower panel of this table contains studies not directly related to the measurement of wage differentials or to the countries under study, yet interesting for the way informality is defined – see the discussion below.

Second, some authors have attempted to deal with the unobserved characteristics that affect both the selection into a particular sector and earnings levels by explicitly introducing selection equations. Arguably, a particular challenge pertains to finding convincing instruments for the selection, i.e., variables that explain location into a particular sector without affecting wages. Studies rarely discuss the relevance of their instruments, however, nor do they conduct sensitivity checks on the choice of instruments.[6] Another issue, related to the first point, is that introducing selection into quantile estimation is not a straightforward procedure and shows some difficulties.[7]

In the present paper, we adopt the usual alternative method to control for unobserved heterogeneity, namely the use of panel data to estimate models with fixed effects (FE). This is also the path followed in Badaoui et al. (2008) and Botelho and Ponczek (2011), two studies closely related to ours. Their analysis is however limited to estimations at the mean and for one country at a time (South Africa and Brazil respectively). For the reasons motivated in the introduction, and more particularly because South Africa is seen as a particular case given its relatively small informal sector, we believe it is important to provide a genuine comparison of different countries with different backgrounds.

Interestingly, the literature has already exploited panel information on workers' moves between formal and informal sectors. This was used to simply calculate the wage changes of those moving across sectors (e.g, Funkhouser, 1997, or Maloney, 1999) or to directly learn from transitions across sectors (Bosch and Maloney, 2007).[8] Our empirical strategy is different and consists in using switches across sectors to capture the informal wage

[6]For instance for Brazil, Carneiro and Henley (2001) and Tannuri-Pianto and Pianto (2008) use establishment size, position in the household, payment methods, other household income, work hours and multiple job holding. Establishment size may indeed indicate informal employment, yet informality cannot be reduced to a small firm effect, as discussed below. A related concern applies to payment methods. Critically, both variables are possibly correlated with wage levels. Other household income and work hours may help to capture the possibility of entering informal employment for secondary workers who require flexible forms of employment for a better work-family balance (Marcouiller et al., 1997). This type of instrument is however not relevant for male salary workers, i.e. the largest group and the focus of our study. The same argument applies to whether the partner is already in the formal sector so that the family is covered by social security (cf., Pratap and Quintin, 2006).

[7]The use of a selection equation and an approximation of the Mills ratio in the quantile estimation is suggested by Buchinsky (1998) and adopted in Tannuri-Pianto and Pianto (2008). As recently shown in Huber and Melly (2011), this approach relies on the assumption of conditional independence between the error terms and the regressors given the selection probability, an assumption which implies that all quantile regression curves are parallel. This naturally limits considerably the usefulness of the quantile regression method – and the heterogeneity often found in estimates across quantiles actually reflects a violation of the conditional independence assumption.

[8]Gong et al. (2004) and Gong and van Soest (2002) also provide evidence on sector transitions and (mean) earnings mobility in Mexico, using dynamic multinomial logit and random effect models.

gap while purging estimations from FE. We describe the general principle in the next sub-section.

Third, standard wage equations are potentially vulnerable to misspecification problems, notably because of the linearity assumption on the covariates and the use of observations outside the common support of individual characteristics for formal and informal workers. Matching techniques, which provide the wage outcomes of formal and informal workers only with comparable observed characteristics, have recently been used by Pratap and Quintin (2006) for the analysis of the sector wage gap in Argentina. To reduce the problem of matching workers on several dimensions, they adopt the method of propensity score matching, as usually recommended. For estimations of mean effects, Smith and Todd (2005) also show that combining difference-in-difference estimations and matching techniques is more robust than traditional cross-section matching estimators, as it allows selection on observables as well as time-invariant selection on unobservables. For this reason, Pratap and Quintin (2006) and Badaoui et al. (2008) use a "difference-in-difference" propensity score matching approach to estimate the mean informal wage gap. The present paper extends this approach to quantile estimations, following the propensity score reweighting technique suggested by Firpo (2007).

2.2 The Evaluation Problem

Measuring the informal sector wage gap would require counterfactual information on the wage obtained by an informally employed worker, were she employed in a similar job in the formal sector. That is, we would like to estimate the pure effect of informality as:

$$E[y_{it}(1)] - E[y_{it}(0)], \tag{1}$$

with $y_{it}(I)$ the earnings of individual i at time t when this person is in informal ($I = 1$) or formal ($I = 0$) employment. We do not dispose of observations on the two possible events (a person is either in formal or informal employment at period t). Usual practice thus consists in comparing the earnings of workers in one and the other sector, controlling for differences in observed characteristics. It is possible to improve on this by using panel information. FE estimations first allow controlling for (time-invariant) unobservables that may determine both selection into the informal sector and wages (e.g., unobserved skills, risk aversion, tastes, etc.). In addition, sector switchers can be used to better approximate the pure treatment effect. For instance, using those moving from formal to informal employment between periods 1 and 2, we can approximate the informality effect (1) using $E[y_{i2}(1)] - E[y_{i1}(0)]$. By analogy with the difference-in-difference approach, we need to adjust for possible trends in earnings over the two periods, which is assumed common to the two sectors. We do so by using "stayers" as a control group, for instance those staying in formal employment. Thus the average treatment effect is approximated by:

$$\{E[y_{i2}(1)] - E[y_{i1}(0)]\} - \{E[y_{i2}(0)] - E[y_{i1}(0)]\}. \tag{2}$$

The identification of conditional wage gaps on inter-sector movers requires some caution however. First, as noted by Card (1996), the use of longitudinal estimators can be highly sensitive to measurement errors: even a small fraction of misclassified workers can lead to large biases if the true rate of mobility between sectors is low (attenuation bias). We show that this rate is actually significant and provide an extensive check for potential measurement errors. We thoroughly examine the nature of sector switchers and pay attention to possible asymmetrical effects (depending on the direction of the move).
Second, a limitation of our approach holds in the necessary assumption that, conditional on FE, sectoral switching is random. In other words, the treatment is assumed to be entirely determined by observables and time-invariant unobservables. Whenever unobservable attributes change over time and their variation is correlated with the variations in earnings and sector choice, the FE estimator is inconsistent. For example, if workers anticipate their relative prospects in both sectors and choose a sector based on those expectations (which are not observed), then the transition from one sector to the other is also endogenous.[9] As noted above, we check hereafter for possible asymmetrical effects depending on the direction of the move. This is important indication since (unobservable) time-specific reasons for transiting between sectors, e.g., productivity shocks, are likely to be different in one or the other direction.

2.3 Defining Informality

There is generally no consensus on how to define informality. Some studies opt for the *productive* view, based on job types or firm size (usually firms of less than five workers). Since own-account workers and the owner of microfirms are counted, this definition overlaps with informal self-employment, which is not the focus of this study. Instead, we adopt the so-called *social security* (or legalistic) view, whereby informality refers to the lack/avoidance of formal registration, taxation and labor regulation as well as the

[9] Accounting for endogenous switching associated with time-varying unobservables would certainly require more information and combining FE estimations with an instrumental variable approach to model selection explicitly (see for instance Harding and Lamarche, 2009). As discussed in the previous subsection, however, convincing instruments are hard to find in the present context. The task is even more complicated since instruments would need to vary over time.

lack of social security protection for workers. These aspects are important for welfare considerations as informal sector workers may experience bad work conditions (e.g., no social protection) at the same time as lower wages. Importantly, this legalistic view also acknowledges the possible presence of informal employment within large firms and, hence, corresponds to a much broader definition of informality than the productive view. This aspect is particularly important – firm size classifications may be picking up the effect of firm size on wages rather than other specific effects of informality – and receive some attention in section 5. In addition, the social security definition allows consistently adjusting wages for taxes/social contributions paid by employees in the formal sector only.

With the data at hand, described in section 3, we can identify informality in a consistent and comparable way for all countries. In Mexico employees have to contribute to the social security agency (IMSS). Similarly, employees in Brazil must hold a labor card (*carteira assinada*), the signing of which guarantees them access to formal labor protection. Therefore those wage employees not registered with the social security agency in Mexico or not holding a signed labor card in Brazil are considered as informal salaried. For South Africa, we follow Badaoui et al. (2008) who rely on questions regarding fringe benefits and other aspects of the job that can be used to identify the sector, in particular questions regarding whether the firm provides medical aid and deducts unemployment insurance contributions. Table 1 shows that these legalistic definitions are in line with what is frequently used in related studies for Brazil (e.g., Tannuri-Pianto and Pianto, 2008), Mexico (Gong et al., 2004, Calderon-Madrid, 1999) and South Africa (e.g., Badaoui et al., 2008) as well as other countries characterized by informal employment.[10]

2.4 Labor Markets in Brazil, South Africa and Mexico

Existing evidence shows that informal labor markets are significant in all three countries under study. In Brazil, Carneiro and Henley (2001) indicate that informal employment, defined as the proportion of salary workers without a signed labor card, represents around 25% of the labor force. Several factors are typically blamed for the presence of an informal sector in Brazil, including taxes (de Paula and Scheinkman, 2006), stringent labor legislations (Barros and Corseuil, 2001) and the enforcement of these regulations (Almeida and Carneiro, 2008). The last two aspects have been particularly reinforced following the

[10]For the reason mentioned above, there is a shift in the literature in favor of the legalistic view (cf. Perry et al., 2006). The productive approach seems nonetheless frequently used in the case of Mexico, simply because the many studies focus on self-employed workers (Maloney, 1999, Gong and van Soest, 2002, Marcouiller et al., 1997). Gong et al. (2004) find some overlap between the three usual definitions but social security coverage corresponds better to the firm size classification than to the job-type definition.

1988 constitutional changes which increased the degree of worker's protection and hence labor costs for firms (see Ulyssea, 2010). We conjecture that firms may try to recoup these legal costs by decreasing formal wages. Also, the reduction in state regulation of collective bargaining has led to increasing rent sharing and insider trade union bargaining power, the consequence of which is not rising open unemployment but a growing displacement of workers into the informal sector (Carneiro and Henley, 1998).

In Mexico, Marcouiller et al. (1997) show that the informal sector represents 31% of total employment when defined according to firm size but more than 43% when the social security definition is used. The informal sector in Mexico has been described as a desirable and voluntary-entry segment of the labor market in many studies (for instance Maloney, 1999, Marcouiller et al., 1997). Yet this statement concerns the self-employed rather than salary workers. As explained in the next section, it seems important to focus on each group separately. In the present study, we question whether informal salary work, the vast majority of informal work in Mexico, also shows unusual patterns or is, on the contrary, similar to informal employment in other Latin American countries like Brazil.

According to Kingdon and Knight (2007), the informal sector represents 24% of the South African labor force in 2003. Many studies, surveyed by these authors, report that labor standards, employment protection legislation and the presence of strong trade unions explain the presence of informality – even if some dynamic segments of the informal labor market also exist (cf. Cichello et al., 2005). Informal salary work alone accounts for 11% of total employment (cf., Badaoui et al., 2008), which is smaller than for other countries – yet the informal sector keeps on growing, as noted by Kingdon and Knight (2004). The reason for a relatively small informal sector is its coexistence with classic unemployment, which represents 29% of the total labor force and is due to higher reservation wages compared to lower income countries. Indeed the unemployed who receive some support from within or beyond the household (social grants) may prefer to remain outside the low-tier informal sector. For those who cannot benefit from these safety nets, however, low-productivity informal jobs can be seen as last resort, underpaid activities. Conditional wage distributions estimated in the present study help to quantify this specific, most deprived group of workers.

3 Data and Measure of the Raw Wage Gap

3.1 Data and Selection

We collect panel datasets with enough information to estimate earnings equation in all three countries, i.e. surveys providing job characteristics, incomes, work duration, demographics and education levels of the workers. For Brazil, we use the Monthly Employment Survey (*Pesquisa Mensal de Emprego*, PME) conducted by the *Instituto Brasileiro de Geografiae Estatistica* (IBGE). This is a monthly household survey on the six largest metropolitan areas of Brazil (i.e., Belo Horizonte, Porto Alegre, Recife, Rio de Janeiro, Salvador and Sao Paulo). Households are interviewed four months in a row and re-interviewed eight months later for another four months. We create a panel with observations that are a year apart, focusing on years 2002 to 2007. For South Africa, we use the labor Force Survey (LFS), a bi-annual rotating panel conducted by Statistics South Africa (Stats SA) and covering all provincial areas. Twenty percent of the sampling units are rotated out of the survey and replaced with a new sample every six months; workers are therefore observed five times at most over a two-and-a-half year period. We use the waves of September 2001 to March 2007. For Mexico, we use the Mexican National Occupation and Employment Survey (ENOE) conducted by the *Instituto Nacional de Estadistica, Geographica e Informatica* (INEGI). This is a quarterly survey where workers are observed at most five times over a five-quarter period. We use data from the first quarter 2005 to the third quarter 2008.

Since households are identified over time but individuals are not, we construct panels of individual workers by linking persons within households over time on the basis of gender, race and age. For the baseline estimates, we select workers that are observed at least twice in the data.[11] We restrict samples to urban male workers aged 15-65 and not engaged in any form of education or training. We focus on men because a large proportion of women in all three countries are not active or are engaged in unpaid work. We select only workers in the private sector, which excludes unpaid family workers (whose implicit earnings are difficult to evaluate) and public sector employees. For the latter, there are indeed important differences in institutional mechanisms regulating wages, both across countries and compared to the private sector. In South Africa (resp. Brazil), whites and asians (resp. asians) are excluded from the sample as they are disproportionately represented in the formal sector.[12]

[11] The attrition resulting from this procedure corresponds to 32% of the initial sample for Brazil, 22% for South Africa and 18% for Mexico. Further work should check for the possibility of non-random attrition that could bias results. An interesting procedure would be, in particular, to derive Manski bounds of the conditional wage gaps in this context. There are, as yet, no simple method to obtain them in a quantile estimation framework and even less so when accounting for fixed effects. For partial identification of quantile treatment effects in the presence of sample selection, see Blundell et al. (2007).

[12] Given that different racial groups may be remunerated in different ways due to specialization or discrimination, including these specific groups which are represented only in one sector would obviously impair the condition of common support between sectors. The racial groups excluded from our sample represent less than 1% of the informal sector.

An important step in the selection is the focus on salary workers only. Self-employed workers form a vastly heterogeneous group, from street vendors to firm owners and professional independent workers, and deserve a specific study. We prefer to focus here on a more homogenous comparison between salary workers in formal and informal sectors, a choice also made in many related papers like Badaoui et al. (2008) or Tannuri-Pianto and Pianto (2008). Arguably, comparing the earnings of dependent and independent workers may be misleading insofar as self-employment income includes returns to risk and capital. Our choice is not restrictive: salary workers also represent the vast majority of informal workers in most countries. Finally, the datasets at hand allow us to define informality in a relatively comparable way for all countries, as described in section 2.

3.2 Wages and Tax Calculations

Real hourly wages are calculated from the gross monthly wages and reported work hours in the primary job. For the sake of comparability between countries and over time, earnings are converted into 2002 international dollars using relevant CPI deflators and PPP adjustment factors drawn from the World Development Indicators. The premium associated with formal sector employment is overestimated if taxes and social contributions paid by registered workers are ignored. Thus we use available information to adjust gross wages in this sector, which is consistent with the chosen definition of formality (see similar adjustment in Badaoui et al., 2008, for South Africa). Adjusting for taxes is sometimes seen as a difficult exercise because of data limitation. We argue that the datasets at hand and the nature of the tax systems in the countries under study allow for a reasonable approximation of the taxes paid on labor income.[13] The tax system is progressive in all three countries but the top marginal tax rates are not very high by international standards in Brazil (27.5%) and Mexico (28%). Systems include a flat rebate in South Africa and a refundable and progressive tax credit for low-wage workers in Mexico. In these two countries, income taxation is purely individualized. In Brazil, taxpayers can also file jointly and benefit from a deduction for each dependent relative (i.e., the spouse, if inactive, and children aged under 22, or 25 if in education). We have used available information on family links for the main adults in the household and assumed that other adults were single. For the latter, we thus potentially overestimate tax liabilities; yet most

[13] Tax rules are summarized in Table A.1 in the online appendix. Detailed descriptions of the tax-benefit systems are available from the South African Revenue Service (http://www.sars.gov.za) and from "Microsimulation models for Latin America" (Urzúa, 2012, ed.) found on: http://idl-bnc.idrc.ca/dspace/bitstream/10625/49851/1/IDL-49851.pdf. The precise description of the imputation process adopted in the present study is available from the authors.

of them are young workers with low wages, and hence likely exempt from tax payment. Another usual limitation to tax calculation is the absence of information concerning capital income, which is therefore excluded from the tax base in our simulations. This should concern only a limited number of people at the very top of the distribution, however. We find that only the top 20% of the gross wage distribution is liable for income tax in all three countries. The effect of taxation on the informal wage gap is discussed below.

3.3 Data Description and the Raw Wage Gap

Table 2 describes the selected samples as homogenously as possible across countries. The selection leaves a sample size of 13,710 men with 27,420 panel observations for Brazil; 9,099 men with 20,052 observations for South Africa; and 100,868 men with 260,878 observations for Mexico. Informal employment as defined above accounts for 15% of total salary work in Brazil, 11% in South Africa and 43% in Mexico. For Brazil in particular, this is lower than the share reported in Section 2 because of the selection. Indeed, women and self-employed, excluded from our final sample, are disproportionately represented in the informal sector.

Table 2 shows that net wages are on average larger in the formal sector in all three countries, with a larger average gap in South Africa. We also estimate the propensity to be in the informal sector using a simple probit model. Estimates and marginal effects are reported in Table A.2 (online appendix). Results point toward a U-shaped relationship between age and informality, that is, the young and the old workers are more likely to be in the informal sector. In South Africa and Mexico, the probability of being formal increases with education. For Brazil, only secondary schooling or higher (i.e., more than 11 years of schooling) guarantees a significantly smaller probability of being informal.

The distribution of the unconditional / raw wage differentials between informal and formal sectors is depicted in panel A of Figure 1 for all three countries. The gap based on gross wages is extremely large in South Africa, around 80% on average and between 60% and 110% along the unconditional wage distribution. In the two other countries, it is smaller but nonetheless substantial (around 30% on average), and more uniformly distributed. When considering *net* wages, we find that progressive income taxation is responsible for slightly decreasing the wage differential across sectors in the top quarter of the distribution of all three countries. In Mexico, the refundable tax credit subsidizes formal sector workers in the first three quarters of the distribution and hence increases the raw penalty faced by informal sector workers in the lower part of the distribution. We only use net wages in the rest of the paper.

4 Econometric Approach

As in early studies (e.g., Marcouiller et al. 1997), we first estimate standard Mincer wage equations including an informal sector dummy which captures the conditional wage gap between sectors. Other covariates comprise standard human capital information (age, age squared and education), individual/household characteristics as reported in Table 2 (race, number of children, marital status, region) and broad industry dummies to control for possible structural differences between formal and informal sectors. Estimations are conducted at the mean (OLS) and at various quantiles (QR) on pooled years data with clustered standard errors. Then we control for (time-invariant) unobserved heterogeneity using the panel dimension of the data. The fixed effects (FE) model is simply written:

$$y_{it} = \alpha_i + \gamma_t + x_{it}\beta + I_{it}\delta + \varepsilon_{it}$$

where $E\left[\varepsilon_{it}\left|\alpha_i, x_{it}, I_{it}\right.\right] = 0$ for all individuals i and periods t. The informal sector dummy I_{it} takes value one if worker i is classified as informal at time t. Vector x_{it} denotes a set of controls, α_i the individual FE and ε_{it} an i.i.d. normally distributed stochastic term accounting for possible measurement error. The coefficient δ, interpreted as a measure of the informal sector wage premium/penalty, is derived from the comparison between stayers and movers, as discussed in section 2. That is, the identification of the conditional wage gap is obtained by measuring wage changes of those moving from formal to informal employment between two periods 1 and 2, adjusted by the wage variation of stayers in the formal sector (control group). Indeed, the coefficient δ corresponds exactly to the treatment effect as measured in equation (2), that is:

$$\delta = E\left[y_{i2} - y_{i1}|I_{i1} = 0, I_{i2} = 1\right] - E\left[y_{i2} - y_{i1}|I_{i1} = 0, I_{i2} = 0\right].$$

Importantly, identification is also obtained by all the other possible permutations between statuses.[14] This approach is standard but one must check that the number of transitions across sectors is large enough for a valid use of the FE estimator. We find that 8%

[14] That is:

$$\begin{aligned}
\delta &= E\left[y_{i2} - y_{i1}|I_{i1} = 0, I_{i2} = 1\right] - E\left[y_{i2} - y_{i1}|I_{i1} = 1, I_{i2} = 1\right] \\
&= E\left[y_{i2} - y_{i1}|I_{i1} = 0, I_{i2} = 0\right] - E\left[y_{i2} - y_{i1}|I_{i1} = 1, I_{i2} = 0\right] \\
&= E\left[y_{i2} - y_{i1}|I_{i1} = 1, I_{i2} = 1\right] - E\left[y_{i2} - y_{i1}|I_{i1} = 1, I_{i2} = 0\right]
\end{aligned}$$

with stayers in the informal sector (lines 1 and 3 above) and those going formal (lines 2 and 3 above). These notations do not account for possible differences in the wage penalty whether it is identified on workers moving from formal to informal sectors or on those moving in the other direction. We nonetheless allow for possible asymmetrical effects in the next section.

of panel observations in Brazil, 12% in South Africa and 24% in Mexico correspond to sector changes, which are reassuring numbers regarding the possibility to identify FE. We provide further robustness checks on the validity of the approach in the next section.

The extension of the standard QR model to longitudinal data goes as follows. For any worker i, we can write the τ^{th} quantile of the y distribution conditionally on observables as:

$$F_{y_{it}}^{-1}(\tau \mid x_{it}) = \alpha_i + \gamma_t(\tau) + x_{it}\beta(\tau) + I_{it}\delta(\tau), \forall \tau \in [0, 1].$$

FE α's have a pure *location* shift effect on the conditional quantiles of the response (i.e., they affect all quantiles in the same way). The first FE-QR technique has been suggested by Koenker (2004) as a direct extension of the standard quantile regression approach. A simpler approach has been recently suggested by Canay (2011). Since individual effects α_i are pure location shifters, they can be estimated in a first step by traditional mean estimations (for instance estimation in first differences), then corrected wages $\widehat{y}_i = y_i - \widehat{\alpha}_i$ are estimated on the other covariates by traditional QR.

5 Empirical Results

5.1 Main Results

For each country, we report the estimated coefficient $\widehat{\delta}$, the (conditional) informal wage gap, at the mean (OLS, long-dashed line) and at different quantiles (QR, solid line) in panel B of Figure 1, together with the 95% confidence intervals (short-dashed lines for OLS and shaded area for QR). In panel C, we depict mean and quantile estimates using FE and FE-QR respectively.[15] We first comment on the series of results consistently found in all countries.

[15] In Table A.3 (online appendix), we also report the wage penalty at the mean, the median and two extreme quantiles as well as the (bootstrapped) standard errors. Because of space limitation, we have not reported the full estimation tables – these are available from the authors. Their findings can be summarized as follows. Returns to education typically increase with the education level. Returns to experience (here proxied by age) generally increase as we move to higher quantiles; the same is true for education with a few exceptions (i.e., at lower education levels in Mexico and for university education in South Africa). Many interpretations are possible: higher ability workers may benefit from higher school quality or obtain higher returns to a given experience/education level. Some country-specific results also appear, for instance regional differences (e.g., workers in Sao Paulo benefit from higher pay) and differences by race in Brazil and South Africa.

Baseline Results. All countries show significant mean informal wage penalties when using linear controls only (OLS), i.e. 62% in South Africa, 19.5% in Mexico and 11.5% in Brazil. This is around 20 (10) points smaller than the average raw wage gaps in South Africa and Brazil (Mexico), pointing to the significant role of "better" observed characteristics among formal employees. That is, observed endowments contribute for a quarter of the unconditional wage gap in South Africa, a third in Mexico and two thirds in Brazil. Large pay differences nonetheless remain after controlling for observables, which may partly be explained by unobserved factors. When accounting for time-invariant heterogeneity (FE), mean informal wage gaps indeed decrease by around a third in all countries, i.e., the informal wage penalty is around 19% in South Africa, 9% in Mexico and 4% in Brazil. That is, unobserved characteristics account for more than half of the raw wage gap in South Africa, a third in Mexico and a quarter in Brazil. Time-invariant unobservables are an important factor behind the apparent informal wage gap, even after controlling for a rich set of characteristics. As always, it is not clear which specific factors are at play in individual effects. What we know is that individual effects captured in FE/FE-QR estimations are time-invariant and cannot be sector-specific (identification would require many more years of observation). Hence they do not correspond to some unobservables like the particular job place the worker is in, for instance. They are rather interpretable in terms of workers' unobserved skills, which may reflect intrinsic talent or unobserved school quality. This interpretation is supported by the positive correlation between observed and unobserved characteristics, i.e., the fact that both observed and unobserved characteristics are poorer among informal salary workers. Other interpretations can be put forward, for instance the fact that formal employees may benefit from more efficient networks. In any case, the remaining wage differential points to the fact that formal jobs provide higher earnings *per se*. In the rest of the paper, we try to interpret this result along the different hypotheses suggested in the introduction (segmentation, bargaining power, firm size or job attributes).

Turning to quantile estimations, we first notice that the conditional wage gap is not even along the distribution. With QR, the overall trend is characterized by smaller penalties at the top of the conditional distribution. After controlling for individual effects (FE-QR), the pattern is surprisingly similar in all three countries: informal wage penalties decrease with conditional quantiles. In other words, the largest penalties are to be found in the left tail of the conditional earnings distribution while penalties tend to disappear at the top. Admittedly, that the heterogeneous informal sector exacerbates wage inequality – or, inversely, that the regulated sector compresses the wage distribution – is in line with intuition. These results are overall significant, i.e., for all countries, FE-QR confidence bounds at the two ends of the distributions are not contained in the confidence interval surrounding the FE coefficients. The top of the South African distribution is an exception (due to the imprecise estimations for this country), yet the penalty at quantile 9 is significantly smaller than the penalty at quantile 1.[16] Hence, quantile estimations reveal important and significant differences along the wage distribution that are not captured by usual estimations at the mean.

Specific Results, Interpretations and Reconciliation with the Literature: South Africa. Our results generalize previous findings for South Africa and help to reconcile the apparently contrasted results in Kingdon and Knight (2007) and Badaoui et al. (2008). The OLS wage penalty of more than 60% is in line with recent results by Kingdon and Knight (2007). However, as noted above, more than half of it is due to unobserved heterogeneity. The very large contribution of unobserved skills is in line with Badaoui et al. (2008), who estimate a FE model for South Africa. Yet, these authors find an average conditional wage gap close to zero while a positive informal wage penalty remains in our case (19%). This difference is likely due to the focus of these authors on the years 2001-2003. In fact, when adding these years to our sample and interacting the sector dummy I_{it} with year dummies (in order to obtain a time-varying wage gap $\widehat{\delta}_t$), we also find a very marginal penalty for the early 2000s, in contrast with a sharp increase in the more recent years. Our estimates also generalize the characterization of the informal wage penalty to the whole conditional distribution, showing that unobserved heterogeneity matters at all conditional quantiles. The shape is interesting, notably the fact that the conditional wage gap is very moderate at the top but substantial at the bottom. First, this pattern is qualitatively similar to those obtained for the two other countries, in contrast with the idea that South Africa is very specific because of a relatively smaller informal sector and the presence of unemployment. Second, our results point to the huge heterogeneity characterizing informal employment in South Africa. On the one hand, the penalty faced by workers in the low-tier informal sector is especially large in this country, which supports the view that the legal context matters, as discussed in section 2. Some workers cannot access formal jobs, nor can they "afford" to remain unemployed. They are likely underpaid compared to workers benefiting from binding minimum wages and stronger bargaining power due to unionization. On the other hand, informal wage penalties are smaller than 20% for the upper half of the conditional distribution and smaller than 10%

[16] Arguably, the distribution of the conditional wage gap appears flatter in the middle of the distribution for Brazil so that only the gaps at the two tails are out of the confidence intervals of estimations at the mean (FE). As suggested by a referee, we have tested differences between two successive quantiles, from .1 to .9 ($Q_{.1} - Q_{.2}, Q_{.2} - Q_{.3}, etc$). Out of these 8 differences, 6 are statistically significant in Brazil, 5 in South Africa and all of them in Mexico. Detailed results are available from the authors.

for the upper quarter, pointing to a large segment of unregistered work which is more comparable to the situation in Latin American countries.

Specific Results, Interpretations and Reconciliation with the Literature: Brazil and Mexico. For Brazil, studies report evidence of significant earnings differentials that may favor the segmentation hypothesis (Tannuri-Pianto and Pianto, 2002, Botelho and Ponczek, 2011). Yet evidence is mixed. Carneiro and Henley (2001) show that for some workers, the informal sector may be a desirable form of employment in Brazil. They find that much of the large informal wage gap can be explained by selection bias and consequently favor the competitive markets hypothesis. Our results actually point to modest wage penalties in the first quarter of the conditional distribution (15% with QR and around 7% when accounting for FE) and small or insignificant penalties higher up. Tannuri-Pianto and Pianto (2002) obtain the same pattern as ours, at least qualitatively. Interestingly, they use an alternative approach based on IV-QR. This similarity is encouraging and more systematic comparisons of the two methods for the same country and the same period should be carried out. Quantitatively, however, conditional gaps are relatively small at all levels in our results, which is more consistent with the competitive view supported by Carneiro and Henley (2001). Differences in magnitude between our estimates and those of Tannuri-Pianto and Pianto (2002) may partly be explained by (i) the fact that wage penalties are not identified on the same group of workers, (ii) the fact that formal wages are not adjusted for the effect of taxes and (iii) some of the limitations discussed in section 2 regarding the instruments used to identify the informality effect (note that the evidence by Carneiro and Henley, 2001, also hinges on the validity of their instruments). More recent work by Botelho and Ponczek (2011) provides a robust approach combining instruments and FE estimations on panel data. They obtain an informal (gross) wage penalty of 8% on average, close to our 4% estimate.

Mexico turns out to be an intermediary case. The conditional wage gap ranges from 15% (32% when ignoring FE) at the bottom of the distribution to zero (5%) at the top. Like in the two other countries, accounting for unobservable skills considerably decreases the extent of the penalty, i.e., workers negatively select into informal employment. Our results indicate that the nature of the Mexican labor market is not fundamentally different from that of Brazil. The literature has pointed to a Mexican specificity only when it comes to self-employed workers, as discussed in section 2. It is characterized by the existence of informal self-employment premia on average (e.g., Marcouiller et al., 1997, Maloney, 1999) and particularly at the top of the conditional distribution (cf., Bargain and Kwenda, 2011). When focusing on salary workers as we do here, an informal wage penalty is found on average, slightly larger than in Brazil in the lower conditional quantiles. This result indicates that the real situation of informal salary workers can be masked by lumping them together with self-employed workers, as done in some studies.

Complementary forces may be at work to explain modest penalties in Latin America. First, stringent employment protection legislation provide formal sector employers with an incentive to employ workers on a temporary basis – so that some of the employees of formal sector firms are likely to be *de facto* informal employees. This could explain why in some segment of the labor market, informal and formal workers are relatively similar – this is especially true in the upper part of the conditional distribution, as further analyzed below. Second, informal wage penalties due to formal sector collective bargaining may partly disappear if firms want to recoup the high costs of formal labor. Increasing skill substituability between formal and informal workers may also decrease wage bargaining power in the formal sector, which is what we likely observe in the upper conditional quantiles.

Combining Propensity Score and Fixed Effects. The preliminary conclusions above are based on a series of assumptions that require proper robustness checks. First, our quantile estimations imposed a linearity assumption on the effect of controls. To better account for different distributions of observables between informal and formal sector workers, we suggest combining FE-QR with a direct matching technique, i.e., a propensity score reweighting. The propensity to be in the informal sector, denoted p, can be estimated by binary models using the set of variables that can potentially influence participation in the informal market and the wage rate. Figure A.1 (online appendix) depicts the distributions of the propensity scores for the formal and informal sector workers respectively. The distributions are fairly different between the two groups, especially in South Africa and Mexico. We follow the suggestion of Firpo (2007) and reweight the observations by the inverse propensity score in quantile estimations. The weights used are $1/p$ and $1/(1-p)$ for informal and formal sector workers respectively. Precisely, we first estimate mean FE models, then estimate traditional QR on the FE-adjusted wages, consistently using weighted observations in each of these steps. Bootstrapped standard errors give confidence intervals at each quantile that we report in panel A of Figure 2. It turns out that relaxing the linearity assumption does not fundamentally affect the pattern found in the baseline estimations.[17] This is reassuring and indicates that the potential

[17] We have simply checked that all observations verify the common support assumption. However, a drawback of the reweighting procedure is that the results may become rather unstable as the propensity gets close to 0 or 1. While the number of observations with $0 < p < .05$ (formal sector) or $.95 < p < 1$ (informal sector) is only 5% in Brazil and 7% in Mexico, it is relatively large in South Africa (20%), leading to a large fanning-out of the confidence interval at the two ends of the distribution.

lack of "common support" between the two sectors is not an issue in our case.

Short Panel Issue and Correlated Random Effect Model. In the case of nonlinear operators like QR, FE are not estimated consistently when the number of time periods is small and this inconsistency is transmitted to the estimators of the other covariates of interest (Koenker, 2004, discusses this incidental parameters problem in the case of QR). We provide two additional results to address this issue. First, we conduct a straightforward check that consists in estimating the model separately on samples of 5-year stayers, 4-year stayers, etc. A systematic relationship between the length of the panel and the results would be worrying. We report results in the first two graphs of panel B in Figure 2, for South Africa and Mexico respectively, i.e., the two countries with more than two periods of observations. Results are very similar to the baseline – note however that the small sample size for South Africa explains some divergence at the two tails of the distribution. Second, we estimate a QR model with correlated random effect (CRE), as suggested by Abrevaya and Dahl (2008). This Chamberlain-Mundlak model does not suffer from the incidental parameters problem. It deals with potential correlation between covariates and unobserved heterogeneity by assuming some restricted dependency, namely that unobservables α_i are linearly correlated with the explanatory variables. Thus, in the simple approach à la Mundlak, unobservables are modeled as the mean values of time-varying covariates (over all periods) plus a normally distributed term. Note that this approach requires balanced panels. Results are presented in the top-right graph of panel B in Figure 2 for Brazil (2-period panel) and in the lower panel B for South Africa (4-period panel) and Mexico (2-period and 4-period panels – we provide only two panel sizes for illustration but consistent findings are found in the other cases). These new estimates are consistent with the FE-QR results, even if the conditional wage gap is now larger. Note that this may simply be due to the restrictive account for individual effects in the CRE-QR. It is nonetheless interesting to see that an alternative estimator, relying on different assumptions and characterized by different limitations (see Canay, 2011), points to similar results.

Alternative Specifications and the Role of Firm Size. We check how results vary with the type of control used in the specification of the wage estimation and notably with the inclusion of firm size. Indeed, one possible explanation for informal wage penalties, mentioned in the introduction but not investigated so far, is the firm size effect (see Badaoui et al., 2010). For this sensitivity analysis, we cannot use FE-QR since time-invariant characteristics are captured in individual effects. Hence, we rely on simple QR on pooled data and begin with a basic model including only standard human capital and individual/household characteristics (X_1). Next, we add a series of additional control variables which are arguably endogenous, i.e. industry / occupation types and firm size (X_2). Our baseline is an intermediate case without firm size. Results are reported in panel A of Figure 3. Adding industry types or occupation types does not change results much (not reported). In contrast, adding firm size decreases the informal wage penalty. This is expected given that firm size is likely to be correlated with our "informal sector" dummy I_{it} if the legalistic definition of informality overlaps to some extent with the firm size classification (see the discussion in section 2.3). Further investigations show that in our data, those classified in the formal sector according to the legalistic approach are most often in firms of more than five employees (86% of them in South Africa up to 95% in Brazil). Yet, the overlap is less perfect for those classified in the informal sector according to the legalistic view: 21% of them are working in *large* firms in South Africa, 32% in Mexico and up to 68% in Brazil (aforementioned studies on Brazil actually report that tax evasion is not limited to small and medium-size enterprises, as is commonly believed). Hence, many informal workers are simply unregistered workers within large firms – this is especially true at the top of the distribution and consistent with our finds that wages are equalized across sectors for the most productive workers.[18] More generally, while firm size may explain some of the wage penalty, the other explanations suggested in our main analysis remain valid.

5.2 Checking for Measurement Errors and Additional Results

We further discuss possible issues related to the method at use. We first cover the usual concerns about the validity of FE estimations. Since the identification relies on sector movers, we check that recorded moves are not imputable to measurement errors but correspond to genuine job changes. We also check that movers are not too specific. For that purpose, we show that moves are not limited to a specific part of the distribution and that results are not driven by the specific nature of *job* movers. Keeping in mind the possibility of non-random selection, and because the (unobserved) reasons for transitioning in one or the other direction may be different, we allow for possibly different wage penalties

[18] To better characterize top wage workers in the informal sector, we run a probit with the binary variable taking a value of one if the worker is in the top quintile of this sector. Among significant coefficients, we find that in all countries, the top paid are more often located in economically active areas (e.g., the Sao Paulo region in Brazil), generally have higher education levels (with the exception of South Africa where they more often hold a vocational degree) and more often hold managerial or administrative positions. The presence of unregistered workers in large firms is especially true at the top in Brazil: around 86% (resp. 53%) of informal workers in the top quintile (resp. lower quintiles) are located in firms with 11 or more employees.

depending on the direction of the move. Finally, we check how the informal wage penalty varies with workers' heterogeneity.

Measurement Errors. Admittedly, inter-sector moves could reflect mere measurement error, i.e., flaws in reporting the correct sector status at certain periods. Hence we first verify whether sector transitions are accompanied by actual job changes. We classify any person who reports less than 6 months of tenure as someone who has started a new job. Since tenure information can be missing (or itself contaminated with reporting error), we also associate job moves with significant changes in firm size and changes in occupation and industry (using the most disaggregated categories available in the data). Of all sector moves, which potentially include several moves per worker over the relevant period, 75% in Brazil, 87% in South Africa and 80% in Mexico are accompanied by a change in *at least* one of these characteristics (detailed results available from the authors). "Unexplained" cases may correspond to reporting errors or to persons who remained in the same job but whose status has been changed by their employer. In both cases, the wage of these workers is expected not to change much over time and this group may actually contribute to finding small informal sector penalties. We rerun the FE-QR estimations after excluding observations corresponding to sector moves unexplained by job moves. Results are reported in the panel B of Figure 3. The penalty becomes larger in the lower quantiles for South Africa and, to a lesser extent, in the middle of the distribution for Brazil. Nonetheless, this variant is reassuringly similar to the baseline and shows that our results are not driven by measurement errors.[19]

Characterizing Sector Movers. We also check that movers are not too specific. First, we verify that transitions across sectors are large enough at quantile levels and are not restricted to certain groups of workers. Figure A.2 (online appendix) depicts the number of movers in and out of the informal sector between two periods, averaged over the different waves of the panel, and expressed as a proportion of the size of base-period informal sector quintiles. It turns out that a substantial number of workers move in both directions and do so at all earnings levels. Transitions are slightly more frequent in the upper quintiles in South Africa and Mexico and occur more often from informal to formal sector (especially

[19] Note that changes in firm size do not fully guarantee that a job change has actually occurred. Yet, if a firm expands dramatically over one year, it may become more at risk of being caught defaulting on stipulated regulation and is therefore more likely to register its workers. At the same time, it may also change its wage setting policy – see Badaoui et al. (2010). Then, a substantial change in the firm structure/size may be treated as a reasonable approximation for job/firm change. We nonetheless include a variant where firm size change is not used as a proxy for job move in Figure 3, panel B.

in Mexico and in lower quintiles in Brazil). Overall, however, they do not seem to be overly concentrated at certain levels of the wage distribution.
Second, we characterize sector movers by running additional probits (dependent variable equals to one if the worker moves). It turns out that movers are not extremely different from the overall selected population in terms of their observed characteristics (pseudo-R2 are around .02 for Brazil, 0.06 for South Africa and .01 for Mexico). Only a few characteristics are significant. Movers seem to be younger, less educated and more often single. This picture applies more systematically to those moving from informal to formal sectors. According to our estimations, however, those moving in this direction are not fundamentally different from workers going in the other direction. Note also that moves occur more frequently within certain industries (e.g., construction and trade in Brazil) and that better wage prospects explain only very partly why people move. Transitions are associated with wage increases for some workers only, mainly grouped in the upper part of the distribution.
Third, sector movers may be specific in as much as they belong to the category of job movers. Indeed, job movers may be very different from job stayers as emphasized by human capital models or job-matching models (cf., Farber, 1999). We first check that job moves are not systematically associated to sector moves. Using a conservative definition of 'job changes' (excluding changes in firm size), we find that a majority actually corresponds to moves within the same sector (71% in Mexico, 80% in South Africa, 87% in Brazil). We also estimate FE-QR *on job movers only*, so that the 'control group' is now only those who change jobs while staying in the same sector, either formal or informal. According to the panel C in Figure 4, results appear not to be fundamentally different from our baseline, indicating that the inclusion of job stayers did not lead to a noticeable bias in the baseline penalty estimation. With a conservative measure of job movers, however, the conditional penalty tends to be larger in the upper part of the South African distribution.

Testing for Asymmetrical Effects. Another aspect of the identification strategy that merits discussion is the assumption that the wage penalty is the same for those that move from the informal sector to the formal sector as it is for those that move in the opposite direction. This is not an issue if all the unobservable heterogeneity is time-invariant, as assumed in the FE estimator. However, the transition could be explained by time-specific shocks on worker/job characteristics as discussed in section 2. With the traditional view that jobs are "better" in the formal sector, one would expect that moves into the informal sector are more often the result of negative shocks so that the penalty identified on workers moving in this direction should be larger. We replicate our results when including only one type of transition at a time. Graphs in the panel D of Figure 4 show that results are not fundamentally asymmetrical: results with one or the other type of transition are not statistically different, with exceptions in the middle of the distributions. In particular, results for South Africa confirm that the informal sector wage penalty is larger when identified on the transitions into informality. This is not verified for Latin American countries.

Between-Group Variation. The FE-QR model simply includes a dummy variable for the informal sector and may be seen as misspecified. While in case of misspecification, least square regressions provide a minimum mean squared error linear approximation to the true functions, Angrist et al. (2006) provide a similar result for QR. Therefore our findings have a meaningful interpretation even if the true informal wage penalty depends on the covariates. Nonetheless, we relax the assumption that returns to education and experience are identical in the two sectors and examine the heterogeneity of the informal wage penalty by simply interacting it with workers' age and education levels. Detailed results, available from the authors, essentially show that younger workers face larger penalties at all percentiles, especially in Brazil and South Africa. This is in line with the traditional view that the informal sector is, for some younger workers, a temporary state where they can gain some training (Bosch and Maloney, 2007). Education levels seem to affect the wage gap only at the two extremes of the distribution. At the top, the informal wage penalty is smaller in all countries – and even turns into a premium in Brazil – for those with higher education. This is in line with results from Arbex et al. (2010) who show that returns to education are higher for informal workers in the top of the conditional distribution. In lower quantiles, we find that a larger penalty occurs for those with higher education in Brazil and Mexico. This possibly reflects that in this part of the distribution, education has a higher return in the formal sector, either because it acts as a signaling device or because this sector is capital-intensive and highly rewarded as a complement to capital inputs. This result and its interpretations are consistent with those in Gong and Van Soest (2002).

6 Concluding Discussion

The present study suggests a very comprehensive empirical approach to estimate informal employment wage penalties along the conditional wage distribution in South Africa, Mexico and Brazil. We adjust for taxes paid in the formal sector, control for time-invariant unobservables using panel information and relax the linearity assumption on observables using propensity score weighting. Applying this empirical approach uniformly to all three

countries, and using comparable definitions of informality, allow performing sound international comparisons. Results are robust to using relatively short panels, to measurement errors and to model misspecification.
We find a series of consistent results for all countries: Estimations along the conditional distribution point to large within-group heterogeneity, with a significant informal wage penalty observed at the bottom of the conditional distribution and disappearing at the top. The informal (formal) sector increases wage dispersion (compression), which is an important aspect of overall wage inequality in developing countries. Workers in the formal sector have both better observed and unobserved skills, and the latter are responsible for a substantial share of the unconditional earnings gap in all countries. The informal wage gap does not boil down to a firm size effect.

Beyond these common features, the informal wage penalty shows different magnitudes in the three countries. Largest penalties are to be found in the lower conditional quantiles of South Africa. While it is hard to judge the extent to which wage penalties compensate for net gains in non-cash attributes attached to informal employment, wage gaps in an order of $25-35\%$, our confidence interval for quantile 1 in South Africa, seem more in line with the traditional view of a low-tier informal sector (Kingdon and Knight, 2007). For this group, policy action is required to limit labor market rigidities, at least those which leave informal workers out of formal activities, and to improve the financial conditions of informal workers. Nonetheless, the fact that the wage penalty is not constant along the distribution suggests that policies aimed to levy labor market regulation should not be applied in a blanket fashion. Arguably, the fact that raw wage gaps are larger than conditional gaps highlights the role of human capital and suggests that the key to more equitable labor markets fundamentally also pertains to additional efforts towards building workers' capabilities. Yet, the prevailing role is played by unobserved abilities so that further research should better define the nature of these skills.
For Brazil and Mexico, informal wage penalties are more modest all along the distribution. Labor market regulations likely explain the extent of informality in these countries and possibly the necessity for firms to recoup some of the labor costs on formal job pay, which could partly explain lower informal wage gaps. While some studies conclude about the presence of dualistic labor markets in Brazil (Botelho and Ponczek, 2011), our estimates provide weak support to this hypothesis. Modest pay gaps, lower than 10%, all along the conditional distribution are more in line with the view of relatively integrated labor markets. This conclusion is shared by the upper half in Mexico. So far, Mexican labor markets were viewed as relatively competitive in studies focusing on the self employed (Maloney, 1999, Marcoullier et al., 1997). These results are also consistent with the fact

that sector mobility is not as restricted as the traditional dualistic view would predict (see Ulyssea, 2010, for Brazil). If any, segmentation in the two Latin American countries would concern only a fraction of workers in the lower conditional quantiles – and other interpretations can possibly explain larger penalties in the lower tail, notably differences in bargaining power across sectors (Carneiro and Henley, 1998). Improving workers' capacities seem also important in these countries – especially in Brazil where it accounts for two-third of the unconditional sector wage gap. More skill substitutability between formal and informal workers would also contribute to discipline formal sector wage bargaining.

Future research should address several limitations. Longer panel with movers changing sector several times could be used to measure the distributional shifts due to individual heterogeneity. Moreover, wage gap measures are only part of a more complete welfare analysis. As Badaoui et al. (2008), we have attempted to account for taxes and social contributions to improve the rending of financial situations in the formal sector. However, we do not know how much of the tax wedge is already absorbed in gross wages, and a more complete model of labor demand and supply would be required to measure tax incidence. Moreover, taxes represent only part of the factors that affect net earnings. Accounting for other cash or non-pecuniary advantages attached to a particular sector represents a considerable challenge but a necessary improvement.

References

[1] Abrevaya, J. and Dahl C.M. 2008. "The effects of birth inputs on birthweight: evidence from quantile estimation on panel data", *Journal of Business and Economic Statistics*, 26: 379-397

[2] Almeida, R. and Carneiro P. 2008. "Enforcement of labor regulation and firm size", *Journal of Comparative Economics*, 37: 28–46

[3] Amuedo-Dorantes, C. 2004. "Determinants and Poverty Implications of Informal Sector Work in Chile", in *Economic Development and Cultural Change*, 52(2): 349-368.

[4] Angrist, J., V. Chernozhukov and Fernández-Val I. 2006. "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure", *Econometrica*, 74(2): 539-563

[5] Arbex , M., Galvao, A.F., Gomes F. 2010. "Heterogeneity in the Returns to Education and Informal Activities", *Economic Inquiry*

[6] Badaoui, E. Strobl, E. and Walsh, F. 2008. "Is There an Informal Employment Wage Penalty? Evidence from South Africa", *Economic Development and Cultural Change*, 56: 683–710

[7] Badaoui, E. Strobl, E. and Walsh, F. 2010. "The Formal Sector Wage Premium and Firm Size", *Journal of Development Economics*, 91(1): 37-47

[8] Bargain, O. and Kwenda P. 2011. "Earnings Structures, Informal Employment and Self-Employment: New Evidence from Brazil, Mexico and South Africa", *Review of Income and Wealth,* 57: S100-S122

[9] Barros, R.P. and Corseuil, C.H. 2001. "The impact of regulations on Brazilian labor market performance", Research Network Working Paper No. R-427, Inter-American Development Bank.

[10] Blundell, R., Gosling, A., Ichimura, I., and Meghir, C. 2007. "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds." *Econometrica* 75: 323-363

[11] Bosch, M. and Maloney W. 2007. "Comparative Analysis of labor Market Dynamics Using Markov Processes: An Application to Informality", World Bank.

[12] Botelho, F. and Ponczek, V. 2011. Segmentation in the Brazilian Labor Market. *Economic Development and Cultural Change*, 59(2): 437-463

[13] Buchinsky, M. 1998. "The dynamics of changes in the female wage distribution in the USA: A quantile regression approach". *Journal of Applied Econometrics*, 13: 1-30.

[14] Calderon-Madrid, A. 1999. "Job Stability and Labour Mobility in Mexico During the 1990s", working paper, Centro de Estudios Economicos, Mexico

[15] Canay, I. A. 2011. "A Simple Approach to Quantile Regression for Panel Data," *Econometrics Journal*, 14 (3): 368-386

[16] Card, D. 1996. "The effect of unions on the structure of wages: A longitudinal analysis," *Econometrica*, 64(4) : 957-79

[17] Carneiro, F.G. and Henley A. 2001. "Modeling formal vs informal employment and earnings: microeconometric evidence for Brazil", *Annals of the XXIX National Meeting of Economics of ANPEC.*

[18] Carneiro, F.G. and Henley, A. 1998. "Wage determination in Brazil: the growth of union bargaining power and informal employment", *Journal of Development Studies*, 34(4): 117-138.

[19] Cichello, P.L., G.S. Fields and Leibbrandt M. 2005. "Earnings and Employment Dynamics for Africans in Post-apartheid South Africa: A Panel Study of KwaZulu-Natal", *Journal of African Economies*, 1-48.

[20] de Paula, A., and Scheinkman J.A. 2009. "Value Added Taxes, Chain Effects and Informality", *Review of Income and Wealth*, 57

[21] Farber, H. 1999. "Mobility and stability: the dynamic of job change", in Ashenfelter and Card (eds), *Handbook of Labor Economis*, vol 3b, 2439-2483

[22] Firpo, S. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects", *Econometrica*, 75(1): 259-276

[23] Funkhouser, E. 1997. "Mobility and labor Market Segmentation: The Urban labor Market in El Salvador", *Economic Development and Cultural Change*, 46(1): 123-153

[24] Gong, X., A. van Soest, and Villagomez E. 2004. "Mobility in the urban labor market: a panel data analysis for Mexico", *Economic Development and Cultural Change*, 53(1): 1-36

[25] Gong, X. and van Soest A. 2002. "Wage differentials and mobility in the urban labor market: a panel data analysis for Mexico", *Labor Economics*, 9(4): 513-529

[26] Harding, M, and Lamarche C. 2009. "Instrumental variable quantile regression estimation for a model with fixed effects", *Economics Letters* 104: 133-135

[27] Huber, M. and Melly B. 2011. "Quantile regression in the presence of sample selection", mimeo.

[28] Kingdon, G. and Knight J. 2004. "Unemployment in South Africa: The Nature of the Beast", *World Development*, 32, 3: 391-408,

[29] Kingdon, G. and Knight J. 2007. "Unemployment in South Africa, 1995-2003: Causes, problems and policies", *Journal of African Economies*, 16 (5): 813-848.

[30] Koenker, R. 2004. "Quantile Regression for Longitudinal Data", *Journal of Multivariate Analysis*, 91: 74-89.

[31] Maloney, W.F. 1999. "Does Informality Imply Segmentation in Urban labor Markets? Evidence from Sectoral Transitions in Mexico", *World Bank Economic Review*, 13(3): 275-302

[32] Marcouiller, D. and V. Ruiz de Castilla and Woodruff C. 1997. "Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru", *Economic Development and Cultural Change*, 45(2): 367-92.

[33] Perry, G., W. Maloney, O. Arias, P. Fajnzylber, A. Mason, Saavedra-Chanduvi J. 2006 (eds.). *Informality: Exit and Exclusion*, World Bank latin America and Caribbean Studies.

[34] Pratap, S. and Quintin E. 2006. "Are labor Markets Segmented in Developing Countries? A Semiparametric Approach", *European Economic Review*, 50: 817-41

[35] Smith, J. and Todd P. 2005. "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125: 305–53.

[36] Tannuri-Pianto, M.E. and Pianto D.M. 2008. "Informal employment in Brazil – A choice at the top and segmentation at the bottom: A Quantile Regression approach", Serie de Textos para Discussao n 236 - Universidade de Brasilia

[37] Ulyssea, G. 2010. "Regulation of entry, labor market institutions and the informal sector". *Journal of Development Economics* 91: 87-9

Table 1: Estimation Techniques and Informality Definition in Related Studies

Study	Country	Informal sector definition	Conditional wage gap estimated at:		Accounting for selection into sectors:		
			mean	quantiles	IV	Panel estimation	Matching or PS weighting
Cichello et al (2005)	South Africa (SA)	productive (2*)	yes#	no	-	-	-
Kingdon and Knight (2004)	South Africa (SA)	social security (a)	yes	no	yes	no	no
Badaoui et al. (2008)	South Africa (SA)	social security (a)	yes	no	no	yes	yes
Maloney (1999)	Mexico (M)	productive (1)	yes	no	no	no	no
Gong and Van Soest (2002)	Mexico (M)	productive (2)	yes	no	@		
Calderon-Madrid (1999)	Mexico (M)	social security (a)	yes##	no	no	no	yes
Carneiro and Henley (2001)	Brazil (B)	social security (b)	yes	no	yes	no	no
Botelho and Ponczek (2011)	Brazil (B)	social security (b)	yes	no $	yes	yes	no
Tannuri-Pianto and Pianto (2008)	Brazil (B)	social security (b)	yes	yes	yes	no	no
The present study	**SA, M and B**	**social security (a,b)**	**yes**	**yes**	**no**	**yes**	**yes**
Funkhouser (1997)	El Salvador (ES)	productive (1, 2, 2*)	yes#	no	no	yes	no
Pratap and Quintin (2006)	Argentina (A)	social security (a,b)	yes	no	no	yes	yes
Marcouiller et al. (1997)	ES, M and Peru	productive (1, 2)	yes	no	yes	no	no
Gong et al. (2004)	Mexico (M)	soc. sec. (a), prod. (1,2)	measure transition across sectors		@	-	-
Bosch and Maloney (2007)	M, B and A	social security (a,b)	measure transition across sectors		-	-	-
Amuedo-Dorantes (2004)	Chile	social security (b)	effect of informality on poverty		yes	no	no

(1) firm size (informal = worker in firm of less than five employees, excluding professionals (ex. doctor))
(2) nature/type of the job (informal = piece-workers and own-account workers, formal = fixed wage, cooperative workers, professionals)
(2*) nature/type of the job (informal = casual wage, in non-professional self-employment or in domestic service)
(a) informal = do not receive fringe benefits / social security coverage, do not pay social security contributions, not registered with the social security agency
(b) informal = not holding a signed labor card
measured using wage changes due transitions across sectors
model search and sector mobility
$ separate estimations on earnings quartiles
@ dynamic multinomial logit panel data model with random effects (equations for sector choice and wages, with exclusion restrictions)

Table 2: Selected Samples: Descriptive Statistics

	Brazil			South Africa			Mexico	
Variable	Formal	Informal		Formal	Informal		Formal	Informal
Gross hourly wage	4.77	3.53		2.54	0.99		2.77	2.30
	(6.94)	(4.92)		(3.67)	(1.60)		(2.33)	(1.96)
Net hourly wage	4.43	3.53		2.39	0.99		2.75	2.30
	(5.52)	(4.92)		(3.07)	(1.60)		(1.91)	(1.96)
Demographics								
Age	36.5	35.9		38.5	38.9		34.6	32.1
# children	3.2	3.3		1.7	2.1		1.8	1.6
household size	3.8	3.9		5.9	6.3		4.6	4.9
% married	0.64	0.54		0.63	0.47		0.62	0.44
Black	0.07	0.07	Black	0.74	0.86			
Brown	0.32	0.33	Coloured	0.26	0.14			
White	0.61	0.60						
Education								
No Schooling	0.01	0.01	No schooling	0.09	0.15	No Schooling	0.02	0.04
1-3 years	0.04	0.04	Primary	0.31	0.40	1-3 years	0.04	0.08
4-7 years	0.24	0.24	Secondary	0.53	0.42	4-7 years	0.24	0.34
8-10 years	0.18	0.18	Vocational	0.07	0.03	8-10 years	0.45	0.40
11+ years	0.53	0.53	University	0.001	0.00	11+ years	0.25	0.13
Province								
Recife	0.06	0.04	Western Cape	0.21	0.11	> 100,000 Inhab.	0.72	0.56
Salvador	0.07	0.06	Eastern Cape	0.09	0.16	15,000-99,999	0.11	0.17
Belo Horizonte	0.16	0.11	Northern Cape	0.08	0.05	2,500-14,999	0.08	0.14
Rio de Janeiro	0.27	0.35	Free State	0.11	0.08	< 2,500	0.08	0.13
Sao Paulo	0.25	0.29	Kwazulu-Natal	0.11	0.14			
Porto Alegre	0.18	0.15	North West	0.11	0.13			
			Gauteng	0.12	0.11			
			Mpumalanga	0.11	0.10			
			Limpopo	0.05	0.12			
Economic sector								
Manufacturing	0.32	0.19		0.36	0.08		0.39	0.20
Construction	0.07	0.15		0.12	0.25		0.11	0.37
Trade & Retail	0.24	0.30		0.28	0.15		0.26	0.16
Services	0.24	0.24		0.17	0.37		0.15	0.09
Transport and Comm.	0.13	0.11		0.07	0.15		0.10	0.18
# panel observations	27,420			20,053			260,878	
# workers	13,710			9,099			100,868	
Share of informal sector	15%			11%			43%	

Statistics concern the selected sample of male aged 15-65, neither in education nor in the public sector. Data covers the period 2002-2007 for Brazil, 2001-2007 for South Africa and 2005-2008 for Mexico. Log hourly wages in 2002 PPP international $. Standard deviations in brackets.

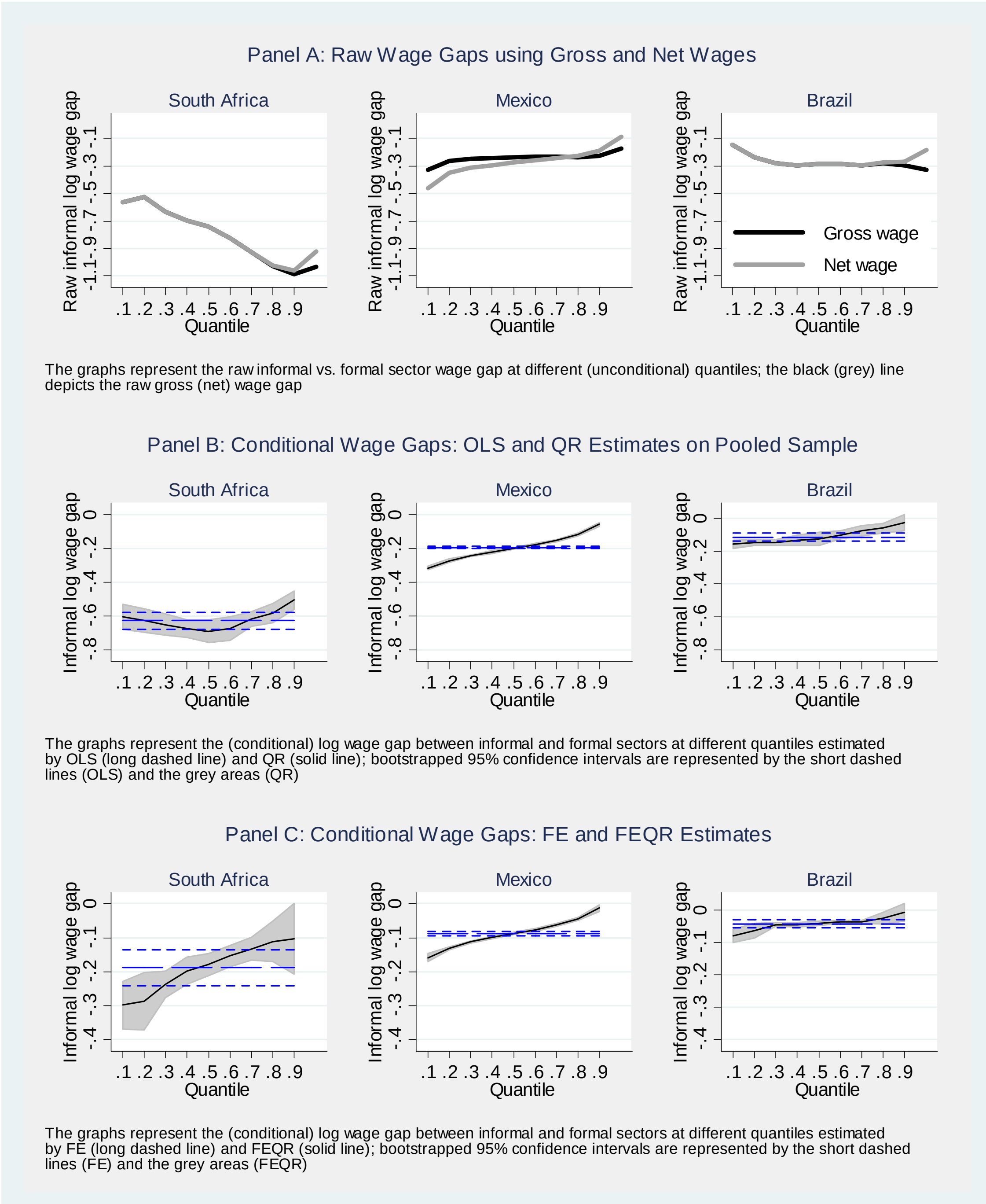

Figure 1: Estimates of the Informal Sector Wage Penalty

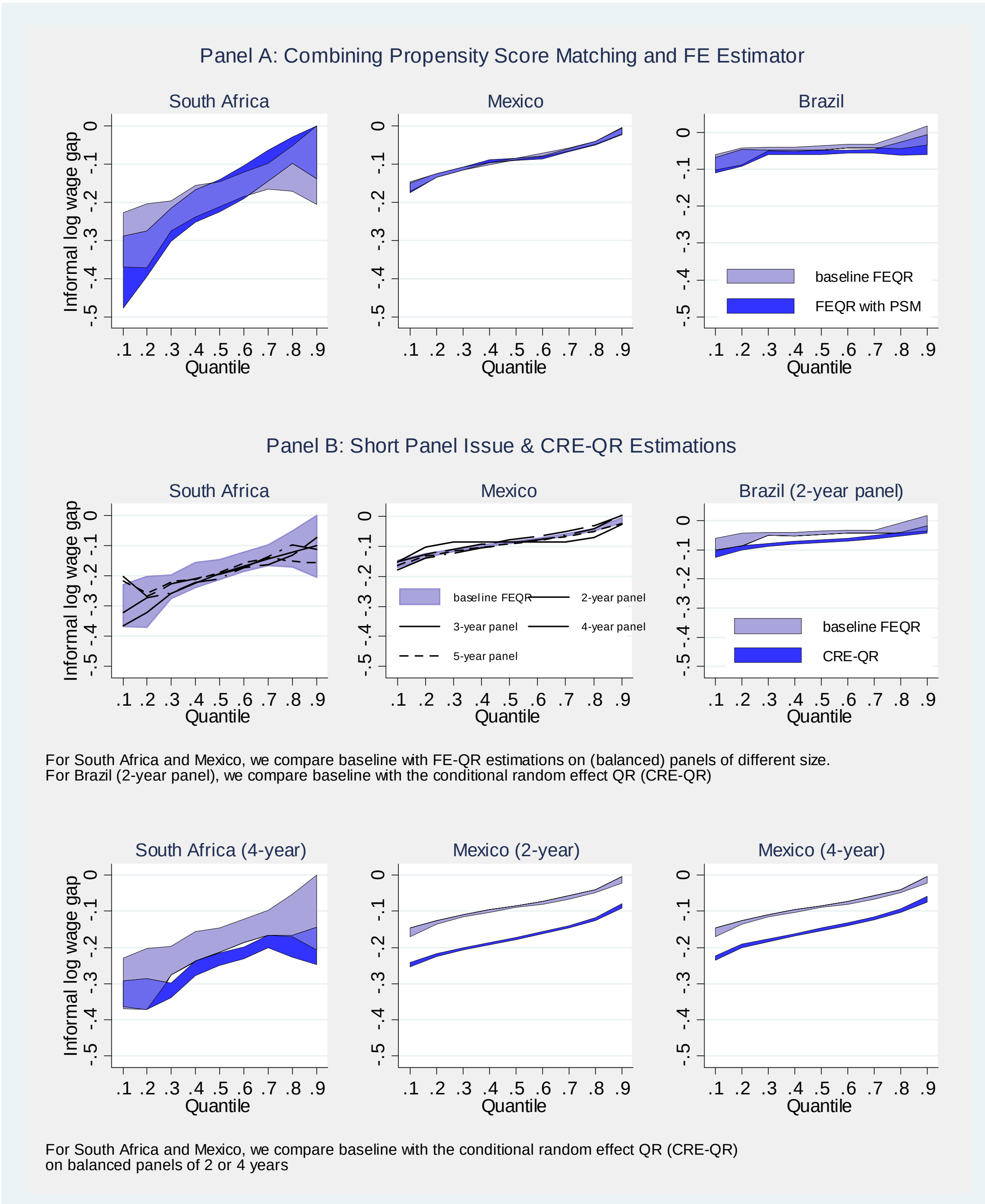

Figure 2: Informal Sector Wage Penalty: Alternative Estimations

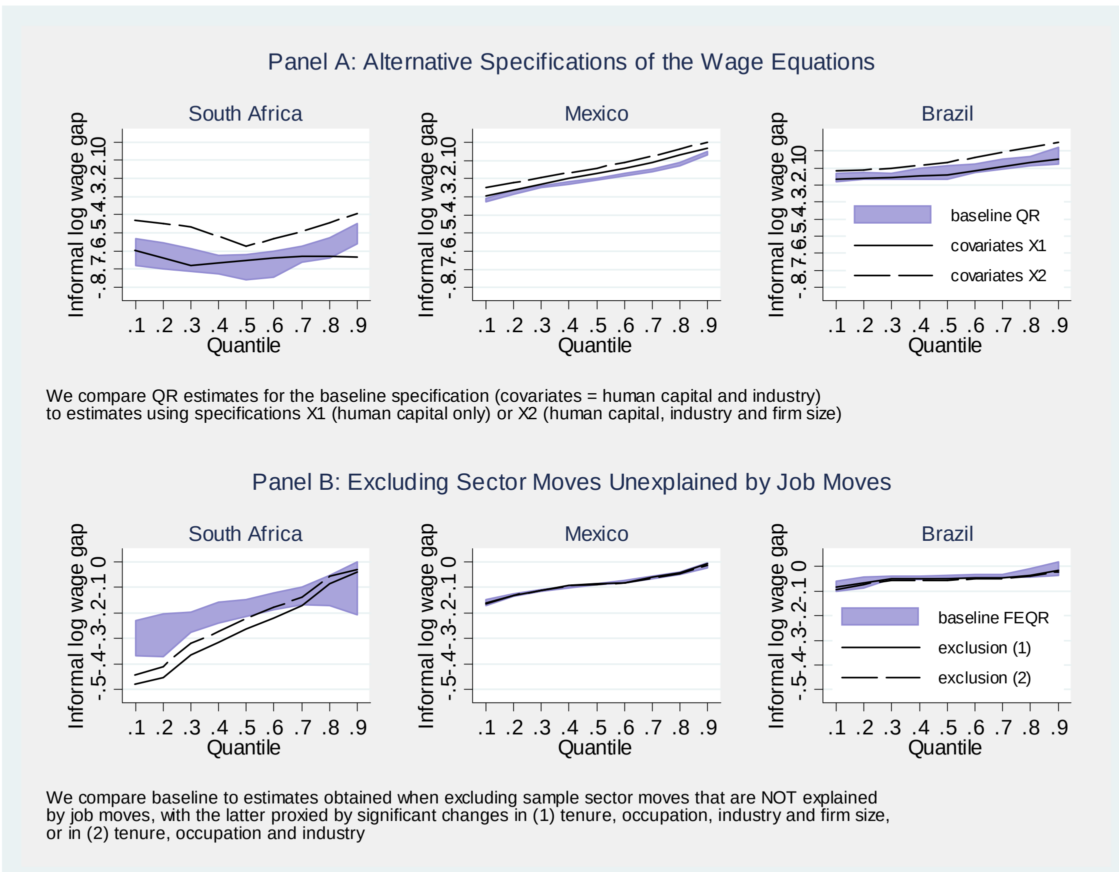

Figure 3: Robustness Checks

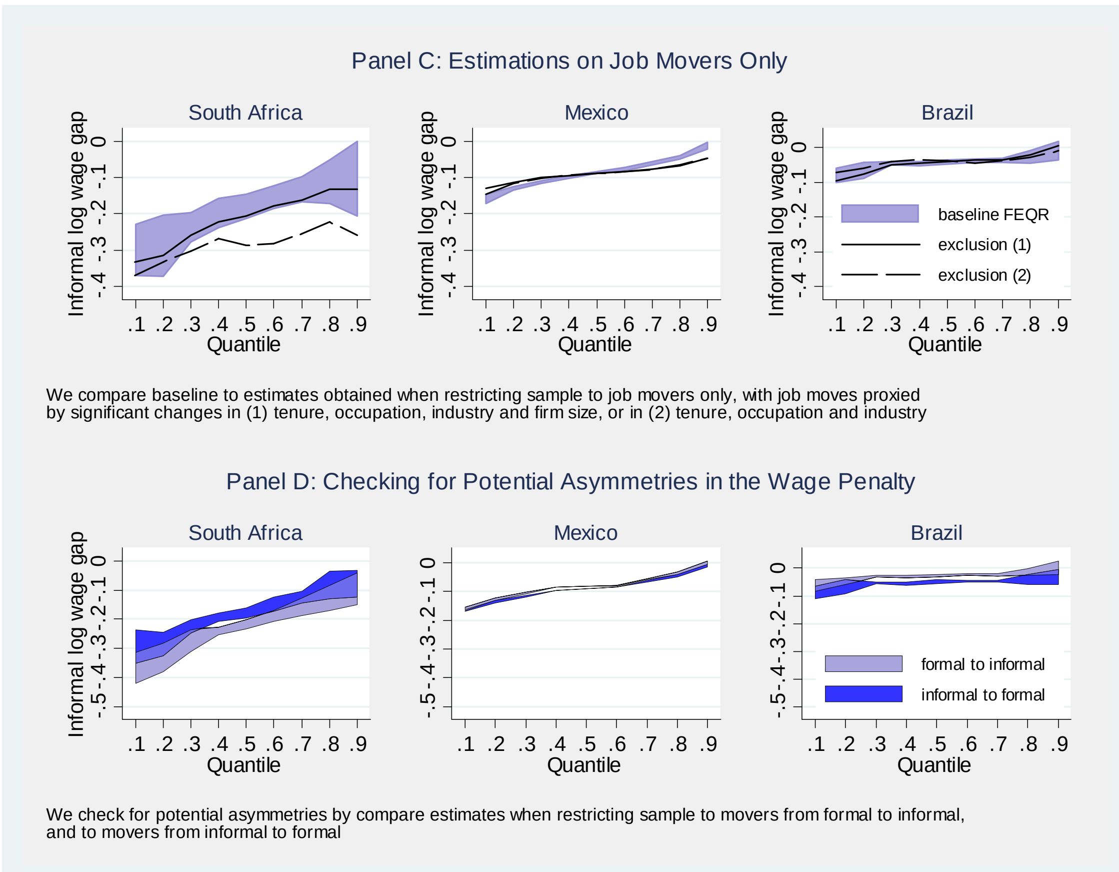

Figure 4: Robustness Checks (cont.)

Online Appendix

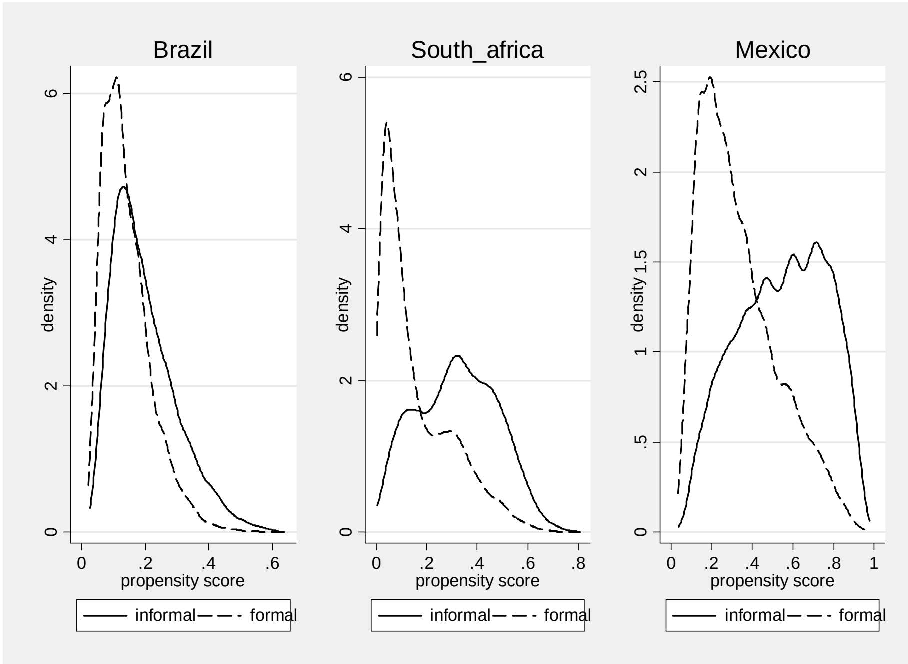

Figure A.1: Kernel-density Estimates of the Propensity Scores

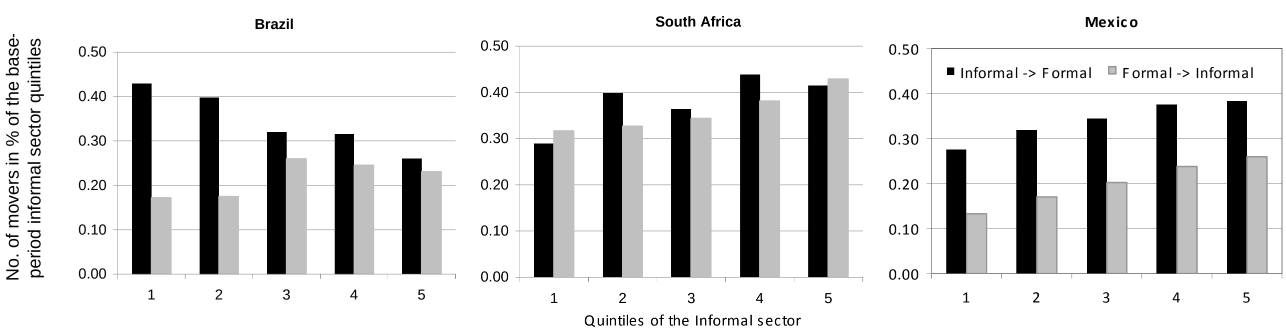

Figure A.2: Distribution of Movers in/out of the Informal Sector

	brackets (annual income)		marginal rate	Others
	in 2002 PPP$	in % of median income		
Brazil	0 … 10,485	0.0 … 1.3	0%	
	10,486 … 20,971	1.3 … 2.6	15%	
	20,971 +	2.6 +	27.5%	
South Africa	0 … 9,091	0.0 … 0.6	0%	A tax rebate of PPP$ 1,636 also applies for all @
	9,091 … 13,468	0.6 … 0.9	18%	
	13,468 … 26,936	0.9 … 1.7	25%	
	26,936 … 37,037	1.7 … 2.4	30%	
	37,037 … 57,239	2.4 … 3.7	35%	
	57,239 … 80,808	3.7 … 5.2	38%	
	80,808 +	5.2 +	40%	
Mexico	0 … 656	0.0 … 0.1	1.9%	People with earnings in the first 3 brackets also receive a refundable tax credit from PPP$ 538 (for zero earnings) down to PPP$ 288 @ #
	656 … 5,570	0.1 … 1.0	6.4%	
	5,570 … 9,789	1.0 … 1.8	10.9%	
	9,789 … 11,379	1.8 … 2.1	16%	
	11,379 … 13,624	2.1 … 2.5	17.9%	
	13,624 … 27,478	2.5 … 5.0	19.9%	
	27,478 … 43,309	5.0 … 7.9	22%	
	43,309 +	7.9 +	28%	

Notes: this table summarizes the tax schedules in force in Brazil, South Africa and Mexico in years 2002, 2002 and 2007 respectively. Our calculations also account for structural changes and nominal adjustments of tax bands occurring at other years. We also account for different treatments of different groups. E.g., for persons aged 65+ in South Africa, there is no second bracket, the theshold to the third is 14,356 and the rebate is increased by 1,010.

@ The upper threshold of the first positive-rate bracket (0.6 and 0.1 times the median income in South Africa and Mexico respectively) is effectively higher (around 1.2 the median income) because of the rebate/tax credit.

 # Hence someone at the end of the 2nd bracket (5,570) has a negative net tax liability of -178; someone close to the end of the 3rd bracket has to pay a net tax of about 500

Table A.1: Tax Schedules

	Variable	Brazil				South Africa				Mexico		
		Coeff.	Std. Err	Marg. Eff.		Coeff.	Std. Err	Marg. Eff.		Coeff.	Std. Err	Marg. Eff.
Demographics	*Ref:*	*white, single*				*black, single*				*Single*		
	Age	-0.094	(0.006)	-0.021		-0.068	(0.015)	-0.008		-0.064	(0.002)	-0.025
	Age squared	0.001	(0.000)	0.000		0.001	(0.000)	0.000		0.001	(0.000)	0.000
	# children	-0.025	(0.014)	-0.005		0.091	(0.021)	0.009		-0.002	(0.002)	-0.001
	household size	0.033	(0.013)	0.007		-0.022	(0.010)	-0.003		0.024	(0.001)	0.009
	Married	-0.200	(0.023)	-0.045		-0.281	(0.061)	0.002		-0.320	(0.006)	-0.125
	Black	-0.072	(0.039)	-0.015	Coloured	-0.246	(0.087)	-0.022				
	Brown	-0.031	(0.024)	-0.007								
Education	*Ref:*	*no schooling*				*no schooling*				*no schooling*		
	1-3 Years	0.134	(0.087)	0.031	Primary	-0.331	(0.080)	-0.029	1-3 Years	-0.153	(0.020)	-0.058
	4-7 Years	0.110	(0.079)	0.025	Secondary	-0.976	(0.087)	-0.096	4-7 Years	-0.320	(0.017)	-0.122
	8-10 Years	0.027	(0.080)	0.006	Vocational	-1.501	(0.143)	-0.074	8-10 Years	-0.580	(0.017)	-0.221
	11+ Years	-0.178	(0.079)	-0.039	University	-1.597	(0.849)	-0.070	11+ Years	-0.797	(0.018)	-0.281
Province	*Ref:*	*Recife*				*Western Cape*				*>100,000 Inhab.*		
	Salvador	0.053	(0.057)	0.012	Eastern Cape	0.938	(0.105)	0.115	5,000-99,999	0.352	(0.008)	0.139
	Belo Horizonte	-0.071	(0.050)	-0.015	Northern Cape	0.203	(0.119)	0.018	2,500-14,999	0.456	(0.009)	0.180
	Rio de Janeiro	0.240	(0.047)	0.056	Free State	0.161	(0.121)	0.016	< 2,500	0.302	(0.009)	0.120
	Sao Paulo	0.289	(0.048)	0.068	Kwazulu-Natal	0.515	(0.113)	0.070				
	Porto Alegre	0.052	-(0.028)	0.009	North West	0.608	(0.115)	0.040				
					Gauteng	0.412	(0.114)	0.022				
					Mpumalanga	0.223	(0.120)	0.133				
					Limpopo	0.994	(0.130)	0.058				
Economic secto	*Ref:*	*Construction*				*Construction*				*Construction*		
	Manufacturing	-0.668	(0.036)	-0.126		-1.521	(0.088)	-0.098		-1.069	(0.008)	-0.371
	Trade & Retail	-0.270	(0.035)	-0.055		-1.041	(0.082)	-0.071		-0.916	(0.009)	-0.315
	Services	-0.352	(0.036)	-0.070		0.075	(0.077)	0.003		-0.846	(0.010)	-0.285
	Transport and Comm	-0.471	(0.041)	-0.085		0.143	(0.100)	0.015		-0.092	(0.010)	-0.035
	Other	0.136	(0.068)	0.032		-1.494	(0.077)	-0.122		-0.283	(0.011)	-0.107
Period	*Ref:*	*year 2002*				*year 2001*				*year 2005*		
	2003	-0.028	(0.046)	-0.006	2002	-0.242	(0.069)	-0.022	2006	0.008	(0.008)	0.003
	2004	0.017	(0.044)	0.004	2003	-0.180	(0.080)	-0.019	2007	0.007	(0.008)	0.003
	2005	-0.007	(0.044)	-0.001	2004	-0.072	(0.090)	-0.009	2008	-0.372	(0.008)	-0.140
	2006	-0.023	(0.043)	-0.005	2005	0.082	(0.091)	0.015				
	2007	-0.077	(0.047)	-0.016	2006	0.115	(0.092)	0.013				
					2007	0.171	(0.093)	0.021				
Constant		1.024	(0.149)			0.685	(0.316)			2.184	(0.032)	

Dependent variable = 1 if informal sector. Standard errors are in brackets.

Table A.2: Probit: Propensity to be in the Informal Sector

Estimation methods		Mean		Q=0.2		Q=0.5		Q=0.8	
		coef.	std.err.	coef.	std.err	coef.	std.err	coef.	std.err
OLS and pooled QR									
	Brazil	-0.115	0.012	-0.148	0.010	-0.126	0.020	-0.060	0.015
	South Africa	-0.627	0.025	-0.627	0.036	-0.689	0.035	-0.582	0.028
	Mexico	-0.195	0.003	-0.274	0.005	-0.201	0.002	-0.120	0.004
FE and FE-QR									
	Brazil	-0.043	0.007	-0.065	0.012	-0.041	0.003	-0.026	0.009
	South Africa	-0.188	0.027	-0.288	0.043	-0.179	0.017	-0.112	0.030
	Mexico	-0.087	0.003	-0.131	0.002	-0.087	0.001	-0.045	0.002

Informal wage penalty = estimated coefficient of the informal sector dummy. All estimations based on the variables reported in the descriptive statistics, except time-invariant characteristics (race, education and region) in the fixed effects estimations.

Table A.3: Informal Wage Gap: Summary of Estimation Results